IMPROVED AUTOMATED LOCALIZATION AND QUANTIFICATION OF PROTEIN MULTIPLEXES VIA MULTISPECTRAL FLUORESCENCE IMAGING IN HETEROGENEOUS BIOPSY SAMPLES

Marina Sapir¹, Faisal M. Khan¹, Yevgen Vengrenyuk¹, Gerardo Fernandez¹, Ricardo Mesa-Tejada¹, Stefan Hamman¹, Mikhail Teverovskiy¹ and Michael J. Donovan¹

¹Aureon Laboratories Inc., 28 Wells Ave, Yonkers, NY 10701, USA

ABSTRACT

We present a novel improvement of our previously published image analysis system for the automated localization and quantification of protein biomarker expression in immunofluorescence (IF) microscopic images. The improvement has been developed primarily for biopsy based images which are by nature of variable quality and heterogeneous. The innovative method is employed for discriminating the biomarker signal from background, where signal may be the expression of multiple biomarkers or counterstains used in IF. The method is dynamic and it derives a threshold for a true biomarker signal based on the relationship between disease and non-disease tissue components. In addition, a new dynamic range feature construction is presented that is less affected by processing and other variations in tissue. The utility of the approach is demonstrated in predicting, based on the diagnostic biopsy tissue, prostate cancer disease progression within eight years after a radical prostatectomy. For this purpose, androgen receptor (AR) and Ki67 biomarker expression in prostate biopsy samples was quantified and features from the proposed approach were shown to be associated with disease progression in a univariate analysis and manifested improved performance over prior systems. Furthermore, AR and Ki67 features were selected in a multivariate model integrating clinical, histological, and biomarker features, proving their independent prognostic value.

Index Terms— Multispectral imaging, biopsy, immunofluorescence microscopy, prostate cancer, prognosis

1. INTRODUCTION

Protein biomarkers are widely used in histopathology for cancer diagnosis, prognosis, and therapeutic response prediction. They provide information on the expression levels of proteins in the cells, allowing for the detection of particular cellular activities associated with the disease state.

One method for protein expression quantitation is IF, where a protein is localized by introducing an antibody labeled with a fluorescent dye into the tissue that binds to the targeted protein [11]. The stained slide is illuminated under a fluorescence microscope with a light source for a specific wavelength. This excitation light is absorbed by the fluorescent dye causing it to emit light of a longer wavelength. The intensity of the emitted light is a measure of the target protein’s concentration. In multiplexing, the tissue is labeled with several antibodies at the same time. The antibodies are labeled by fluorescent dyes with distinct spectral characteristics. Separation of multiple biomarkers is accomplished via multispectral imaging of the tissue followed by spectral unmixing to yield images that represent the expression of the individual antibodies.

Conventionally, IF images are interpreted by pathologists based on the perceived intensity levels of the biomarkers in the objects of interest (e.g. nuclei). This is labor-intensive and suffers from intra- and inter-observer variability. The development of automated systems enables a low-cost, objective alternative to visual scoring of IF.

Quantitation of a biomarker is achieved in two stages. First, a biomarker relevant compartment relevant is detected. Then, the signal is separated from the background within the compartment. The tasks are often accomplished via intensity thresholding. Interactive thresholds [1, 6, 8, 10] are usually followed by computer measurements to quantify biomarker expression. Gordon et al. [6] identified proteins in single-cell images as sets of contiguous pixels above a threshold three standard deviations from the background. Mode of the image histogram together with interactive thresholding is used for identifying background pixels in the AQUA system [1]. Rao et al [9] utilize manual epithelial nuclei delineation to mark the compartment of interest. Limited reproducibility of interactive methods and the poor performance of histogram thresholding on images with low signal-to-background ratios are the primary disadvantages of these existing systems.

We previously presented an automated and reproducible system [10] for biomarker localization and quantification that addressed the above limitations. However, the system required a (relatively small) manually tuned expert gold standard training set in order to train the automated system. Additionally, it was developed on images from TMAs where processing variations were...
minimized and image variability was mitigated. In a diverse biopsy cohort, it tended to over express biomarker positivity.

In this paper we propose an improved method designed to quantitate IF biomarkers in heterogeneous biopsy tissue. The method presented is fully supervised; outcome information is used for training only and it does not require a manual gold standard. Additionally, only the properties of the actual biomarker, in disease and non-disease tissue structures, are used for dynamic thresholding thereby limiting the impact of processing and variation on the other background characteristics employed in our original system.

The improved approach has an increased univariate association with disease-specific outcome and the novel modification increases the ability to quantify a biomarker signal within the relevant subcellular compartments and enhances the robustness of the system against noise variations. Furthermore, the system is extended to quantitate another prognostic biomarker for prostate cancer, Ki-67.

2. IF MULTIPLEX IMAGE ANALYSIS PLATFORM

The underlying platform was the IF multiplex assay we previously presented [10]. This assay consists of nuclear counterstain 4'-6-diamidino-2-phenylindole (DAPI) along with cytokeratin 18 (CK18). There are 2 multiplexes, one with androgen receptor (AR) and another with Ki-67.

The quantification system is designed using the Definiens Enterprise Image Intelligence Suite™ [4]. Valid biological objects (e.g. nuclei) are segmented based on an object-oriented paradigm. Objects, rather than pixels are the smallest units for processing and feature calculation.

It is important in the prognostic cancer prediction to distinguish epithelial nuclei from other tissue components, as that is the relevant compartment for both AR and Ki-67. As described in [10], nuclear objects are segmented and then separated using a colocalization scheme into epithelial (DAPI+/CK18+) and stroma nuclei (DAPI+/CK18-).

3. IMPROVED PROGNOSTIC BIOMARKER EXPRESSION AND QUANTIFICATION

Following the identification of epithelial and stroma nuclei, the next step is to discriminate true biomarker signal from background. Background consists of autofluorescence and non-specific binding of the fluorescent dye to the tissue. After positive object identification, the final step is biomarker quantification where properties such as area and intensity are measured from the classified objects based on the colocalized biomarkers. The positive/negative threshold identification of the prognostic biomarkers and features constructed from them is where the innovative improvements are presented in this paper.

The core issue is that the measured intensity of a biomarker is not only associated with the target protein, but also with nonspecific binding, which may be stronger than the specific binding. The non-specific binding is a result of tissue history (e.g. extraction, processing and storage) and constitutes the “noise” in the image. TMAs are often constructed in standardized protocols intended to mitigate these effects, they are aggravated when working with tissue such as from a biopsy, where variability is substantial.

In [10] we trained a model using background image characteristics to fit a manually trained threshold. We now propose a system where a threshold is selected from percentiles of biomarker expression in stroma nuclei. For example, the 90th percentile of stroma nuclei for AR would be the level of AR expression that 90% of stroma nuclei are below. Using percentiles serves as an inherent dynamic approach as the 90th percentile value will self-adjust depending on the overall expression within an image. In every image, the value for each stroma nuclei percentile is used as a potential threshold in epithelial nuclei. Nuclei with values above this threshold are determined to be positive, and areas and average intensity features per percentile can be calculated; resulting in one area and intensity feature for each calculated percentile. Then the correlation of each of these features is measured univariately with outcome. The stroma nuclei percentile whose features have the best correlation is used as the thresholding value. The pseudocode for this algorithm is presented below:

I. Improved Biomarker Thresholding Algorithm

1) Calculate percentiles of biomarker expression in stroma nuclei
2) For all stroma nuclei percentiles
3) For all images
4) use the value of this percentiles as a threshold for epithelial nuclei
5) calculate intensity and area features in epithelial nuclei based on this threshold
6) end for all images
7) calculate univariate correlation of features with outcome
8) end for all stroma nuclei percentiles
9) the stroma nuclei percentile with the best univariate correlation is used to dynamically determine
10) the biomarker threshold for denovo images
11) correlation
12) the biomarker threshold for denovo images
13) the biomarker threshold for denovo images

The approach is powerful for determining the appropriate threshold given the underlying biology of each biomarker. For example, AR is widely expressed, and only the highest levels are considered predictive. Consequently, the 95th stroma nuclei percentile values for AR were identified as the best threshold. Conversely, Ki-67 is a rarely expressed biomarker and expression is easily identifiable even visually, hence the Ki-67 5th stroma nuclei percentile was identified as the best threshold. This is accomplished without the need for a biomarker threshold gold standard training set, resulting in a fully supervised and automatic method.
If desired, the results can be manually refined given application and biological specific criteria. For instance, the difference in univariate correlation was insignificant for Ki-67 stroma nuclei values in the 5th to 30th percentiles. Hence, as a practical matter, the 30th percentile could be used as a guard against unexpected noise and background expression, since the predictive power of the epithelial nuclei feature would not change. This could also be implemented into the algorithm implementation, to choose the highest percentile of expression that does not degrade performance.

### 4.1. Improved Biomarker Quantification

In addition to biomarker intensity and area, the percentiles permit the calculation of new dynamic relative range features that measures the range of expression through a self-adjusting mechanism. For example, the difference in high and low values of epithelial nuclei percentiles could be assessed and normalized by either (1) the biomarker threshold (a single stroma nuclei percentile) or (2) the same difference in stroma nuclei percentiles:

\[
\frac{EN^{90th} - EN^{10th}}{EN^{90th} - EN^{10th}} \quad (1) \quad \frac{SN^{90th} - SN^{10th}}{SN^{90th} - SN^{10th}} \quad (2)
\]

This relative rise of the biomarker in the epithelial disease-specific components to the stromal non-disease components is a powerful feature to measure the dynamic range of biomarker expression in an image.

### 5. METHODS AND RESULTS

We demonstrate the utility of the proposed improvement to our IF image analysis system in predicting significant prostate cancer disease progression (including metastasis and death-of-disease) in a multi-institutional cohort from the Mayo Clinic, Duke-Durham Veterans Affairs Medical Center, University of Connecticut Health Science Center, and the University Hospital at Uppsala.

Formalin fixed, paraffin embedded biopsy tissue samples for patients treated with radical prostatectomy (RP) between 1989 and 2003 for localized or locally advanced prostate cancer (cT1c-cT3) were studied. The samples were labeled with the DAPI counterstain, and the CK18, AR and Ki-67 biomarkers, and were imaged using the CRI Nuance™ multispectral imaging system [3]. For each sample, 12-bit 1280x1024-pixel grayscale images were acquired at the emission peak wavelength of the DAPI counterstain. For each of the biomarkers, an image stack was acquired for a wavelength range encompassing the emission spectrum of the corresponding fluorescent dye. The resulting image stacks were unmixed using the Nuance™ system to obtain images for each biomarker.

Based on demographic characteristics, 461 patients were stratified into a training set and 230 patients were stratified into a test set. In the training set, similar to [10] a subset of multiplex images were used to train expert thresholds for AR and Ki-67 by a certified pathologist. Threshold models as described in [10] were constructed. Relative positive area, intensity and dynamic relative range features were calculated based on the original system. Additionally, the features were calculated in the training set using the novel methodology described in this manuscript. Both approaches were then evaluated in the test set.

As manifested in Table 1, the proposed method of IF quantification yielded features that were not only more accurate in training, but did not overfit to the outcome and maintained the improved performance in the test set. Additionally, as illustrated in Figures 1 and 2, the new system reduces the over-expression of biomarker positivity in biopsies as observed in the original approach.

As manifested in Table 1, the proposed method of IF quantification yielded features that were not only more accurate in training, but did not overfit to the outcome and maintained the improved performance in the test set. Additionally, as illustrated in Figures 1 and 2, the new system reduces the over-expression of biomarker positivity in biopsies as observed in the original approach.

We also trained multivariate post-RP models based on the biopsy sample in the context of the systems pathology paradigm. In this paradigm, disparate information from clinical (e.g. age), histological (via morphological image features from hematoxylin and eosin-stained tissue) and molecular (IF) profiles are combined in supervised learning to predict cancer progression [5]. Multivariate models were trained for two different endpoints, and both had the relative dynamic range of AR to stroma nuclei ((2) above) and the relative area of Ki-67 positive epithelial nuclei.

Additionally, in results not presented here due to space restrictions, we tested the robustness of feature (2) above, by running 20 samples through two different systems known to have variations in IF processing and observed the resulting relative rise feature. Despite significant differences in the raw epithelial and stromal nuclei percentile values, the feature yielded similar values, indicating robustness.
6. CONCLUSIONS

We propose a novel image analysis system for the automated localization and quantification of protein biomarker expression in IF multiplexed microscopic images that improves upon our original approach. The system is fully automatic and reproducible, precludes the need for manual training, and is robust in heterogeneous biopsy tissue samples. A major component is a novel method for dynamically discriminating true biomarker signal from background through the use of percentiles and supervised correlation with outcome. In the present communication, we demonstrate this system’s application to IF multiplex images of prostate biopsy samples. Features including the area and intensity of AR and Ki-67 along with a novel dynamic range feature construction were shown to have significant univariate correlation with outcome. Additionally, the novel and robust AR dynamic range and Ki-67 relative area features were selected as informative in competition with other predictive factors in a multivariate environment, thereby validating the continued utility of the improved system in practical applications. Our future work will be to reevaluate the thresholds for DAPI and CK18.

7. ACKNOWLEDGMENTS

We would like to thank our colleagues and collaborating partners for their support. US and international patents have been applied for the work presented.

8. REFERENCES